

DriveMLM: aligning multi-modal large language models with behavioral planning states for autonomous driving

Year: 2023 | Citations: 204 | Authors: Wenhai Wang, Jiangwei Xie, ChuanYang Hu

Abstract

Large language models (LLMs) have opened up new possibilities for intelligent agents, endowing them with human-like thinking and cognitive abilities. In this work, we delve into the potential of large language models (LLMs) in autonomous driving (AD). We introduce DriveMLM, an LLM-based AD framework that can perform close-loop autonomous driving in realistic simulators. To this end, (1) we bridge the gap between the language decisions and the vehicle control commands by standardizing the decision states according to the off-the-shelf motion planning module. (2) We employ a multimodal LLM (MLLM) to model the behavior planning module of a module AD system, which uses driving rules, user commands, and inputs from various sensors (e.g., camera, LiDAR) as input and makes driving decisions and provide explanations. This model can plug-and-play in existing AD systems such as Autopilot and Apollo for close-loop driving. (3) We design an effective data engine to collect a dataset that includes decision state and corresponding explanation annotation for model training and evaluation. We conduct extensive experiments and show that replacing the decision-making modules of the Autopilot and Apollo with DriveMLM resulted in significant improvements of 3.2 and 4.7 points on the CARLA Town05 Long, respectively, demonstrating the effectiveness of our model. We hope this work can serve as a baseline for autonomous driving with LLMs.